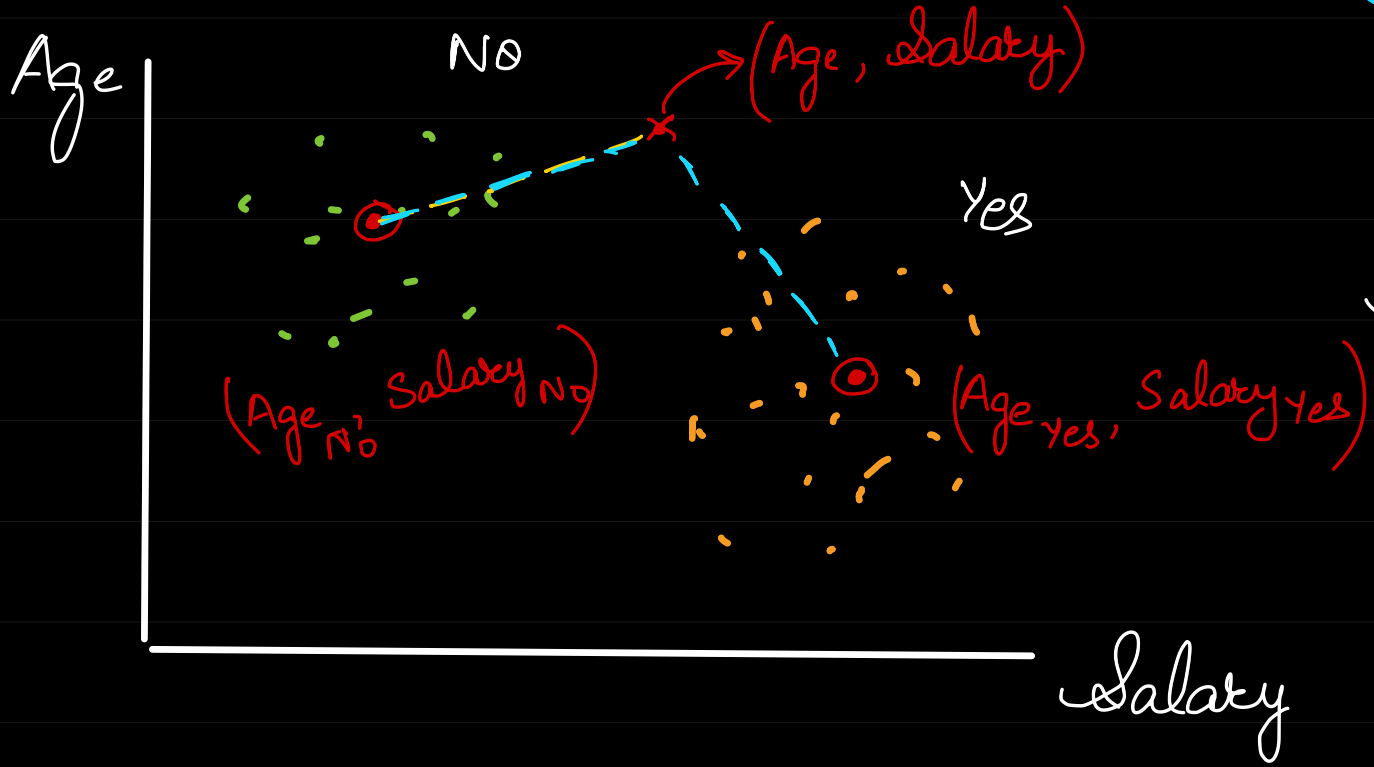
Data Scaling



$$\sqrt{(y_2-y_1)^2+(x_2-x_1)^2}$$

Dist from class No

$$\sqrt{(Age - Age_{No})^2 + (Salary - S_{No})^2}$$

$$\sqrt{70^2 + (4000)^2}$$

Distance from class Yes

$$\sqrt{(Age - Age_{Yes})^2 + (S - S_{Yes})^2}$$

$$\sqrt{80^2 + (2000)^2}$$

Age : 20 - 90 years

Salary : \$ 1000 - \$ 5000

$(90-20)^2$
$(70)^2$

$(5000-1000)^2$
$(4000)^2$

Age	Salary	should we give project
20	1000	No
30	1200	Yes
50	4000	Yes
60	4050	No

Data Scaling

- Min Max Scaling
- Standard Scaling

Min-Max Scaler $[0, 1]$

$$X_{scaled} = \frac{X_i - min(x)}{max(x) - min(x)}$$

$$Age_{scaled} = \frac{Age_i - min(Age)}{max(Age) - min(Age)}$$

$$Salary_{scaled} = \frac{S_i - min(S)}{max(S) - min(S)}$$

$$\frac{min(x) - min(x)}{max - min} = 0$$

1

Standard Scaling / Normalisation

$$X_{scaled} = \frac{X_i - X.mean()}{X.std()}$$

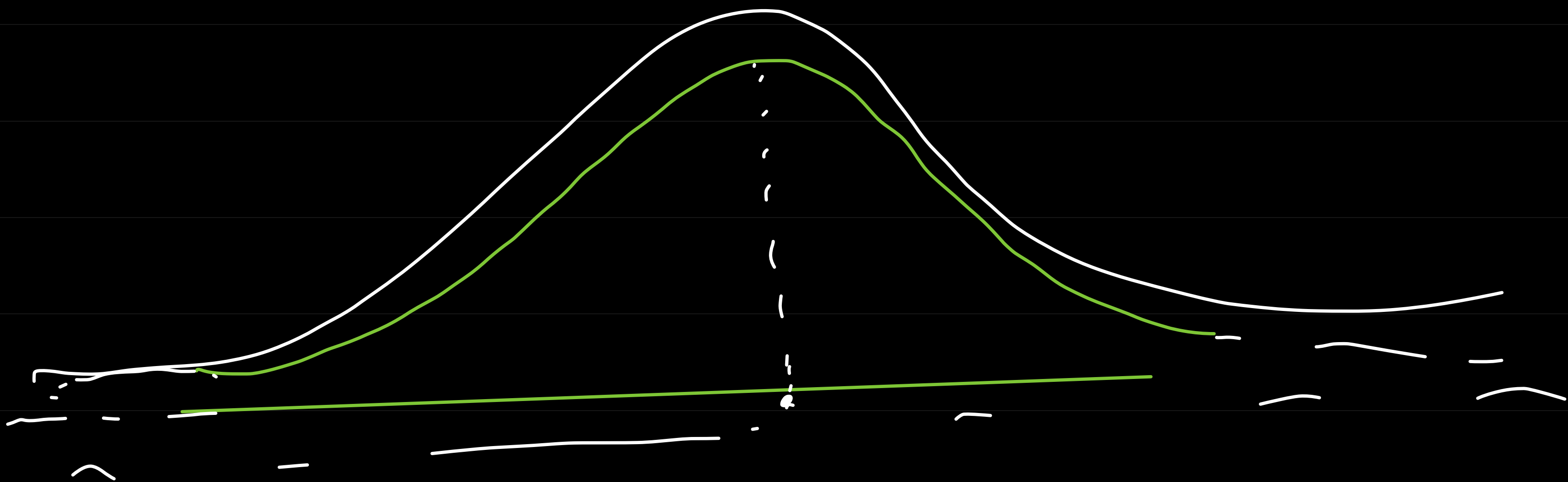
$\mu = 0$
$\sigma = 1$

$$\text{Variance} = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$$

Mean : μ
: $E[X]$
: $\bar{x}$

$$\begin{aligned}
\text{Variance} &= E\left[(X - E[X])^2\right] \\
&= E\left[X^2 - 2X E[X] + E[X]^2\right] \\
&= E[X^2] - 2E[X]E[X] + E[X]^2 \\
&= E[X^2] - E[X]^2
\end{aligned}$$

Standard Deviation (σ) = $\sqrt{\text{Variance}}$



Data Splitting

input

train_input

output

train_output

Outlook	Temp	Humidity	Windy	Play Golf
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Overcast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

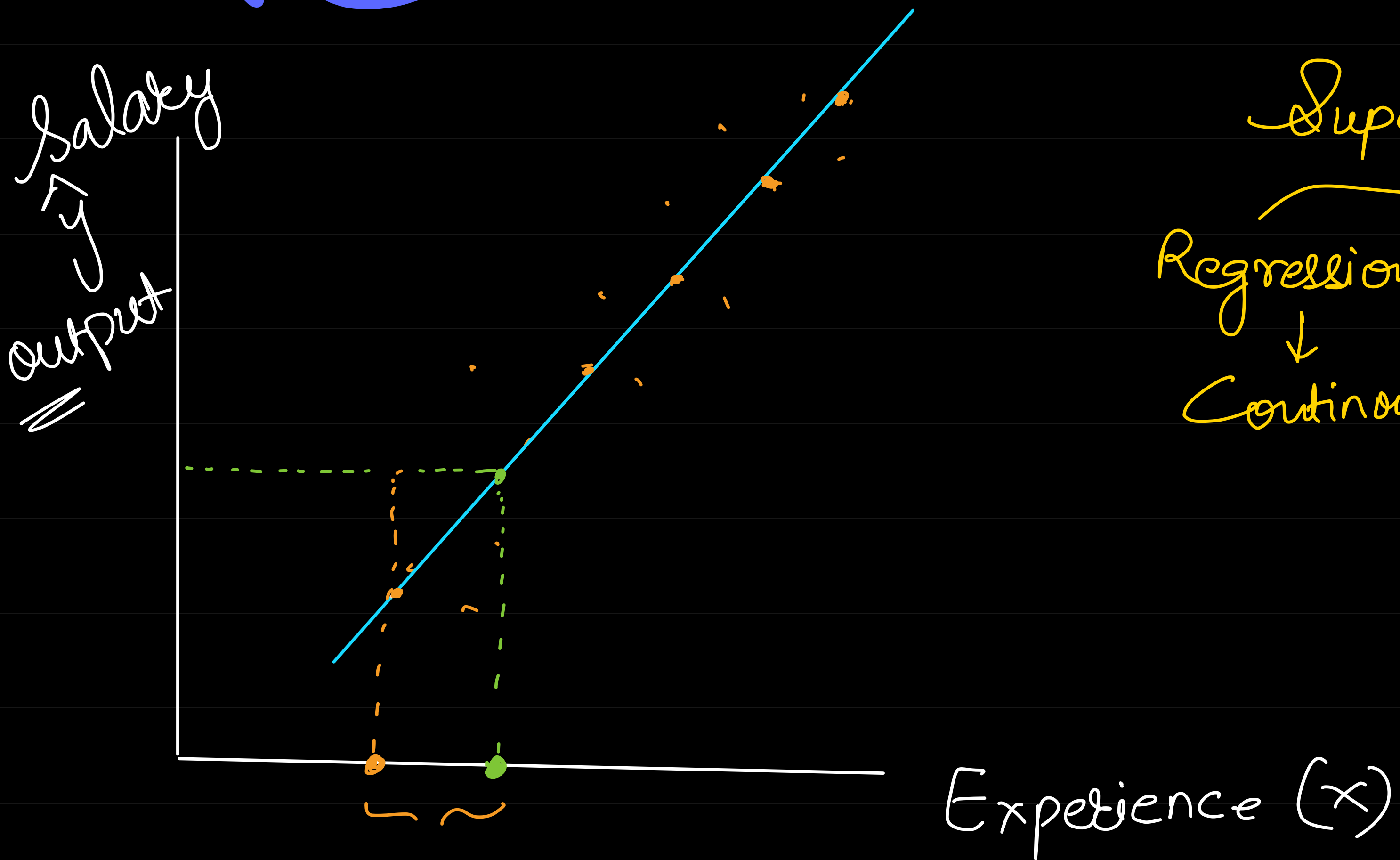
P
P
P
P

test_output.

test_input

predn

LINEAR REGRESSION



Supervised Learning

- Regression → Continous
- Classification → Discrete

$$\underset{\text{slope} \nwarrow}{} \quad \hat{y} = m\,\underset{\uparrow \text{ input}}{x} + c \quad \nearrow \text{intercept}$$

y

$\left.\begin{matrix} m \\ c \end{matrix}\right\}$ weights

$\Rightarrow$ optimal weights.

Step 1: Randomly initialize 'm' and 'c' values.

Step 2: Calculate Cost, error

$$\text{Cost} = \frac{(\hat{y} - y)^2}{2\underline{\underline{n}}}$$

$\underline{\underline{n}} \rightarrow$ no. of instances

y actual train output

$$\Big(\underbrace{m \cdot \overset{x}{\overset{\downarrow}{2}} + \underline{\underline{c}}}_{\hat{y}} - \underset{y}{\cancel{5}}\Big)^2$$

m^2 c^2

Gradient Descent Algorithm

$$m' = m - \alpha \frac{\partial Cost}{\partial m}$$

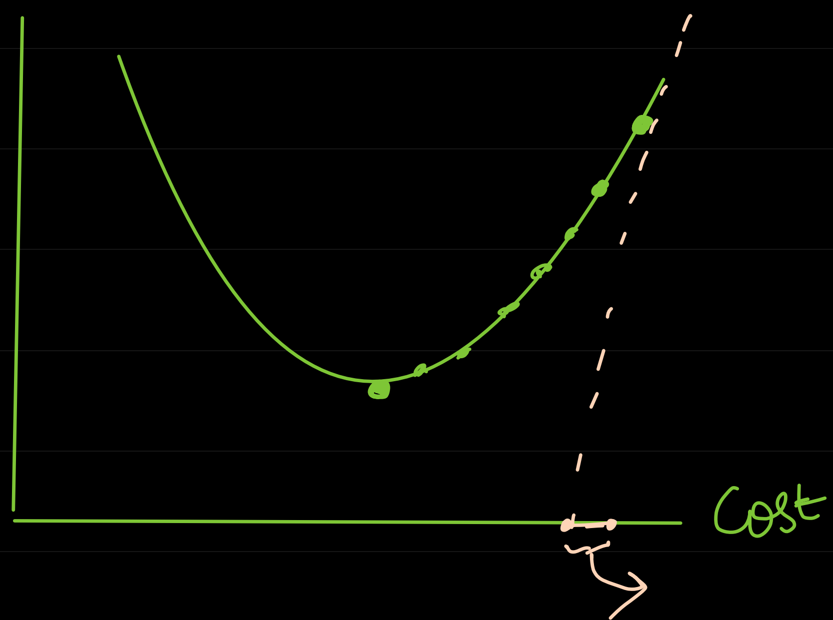
α : learning_rate

0.1

0.2

0.001

$$c' = c - \alpha \frac{\partial Cost}{\partial c}$$



$$\hat{y} = f(x) = mx + c$$

$$\text{Cost} = \frac{1}{2n}(\hat{y} - y)^2$$

Gradient Descent

$$m' = m - \alpha \frac{\partial \text{Cost}}{\partial m}$$

$$c' = c - \alpha \frac{\partial \text{Cost}}{\partial c}$$

$$\frac{\partial Cost}{\partial m} = \frac{\partial Cost}{\partial f} \cdot \frac{\partial f}{\partial m}$$

$$\frac{\partial Cost}{\partial c} = \frac{\partial Cost}{\partial f} \cdot \frac{\partial f}{\partial c}$$

$$df = \frac{\partial Cost}{\partial f} = \frac{\frac{1}{2n}(\hat{y}-y)^2}{\partial \hat{y}}$$

$$= \frac{1}{n}(\hat{y}-y) \cdot \frac{\partial(\hat{y}-y)}{\partial \hat{y}}$$

$$= \frac{1}{n}(\hat{y}-y) \cdot 1$$

$$df = \frac{\hat{y} - y}{n}$$

dm $$\frac{\partial f}{\partial m} = \frac{\partial \hat{y}}{\partial m} = \frac{\partial (mx + c)}{\partial m} = x$$

dc $$\frac{\partial f}{\partial c} = \frac{\partial \hat{y}}{\partial c} = \frac{\partial (mx + c)}{\partial c} = 1$$

$$m' = m - \alpha \frac{\hat{y} - y}{n} \cdot x$$

$$c' = c - \alpha \frac{\hat{y} - y}{n} \cdot 1$$